UDC: 004.93 DOI: https://doi.org/10.15407/jai2021.01.032

DEEP NEURAL NETWORK BASED ON GENERALIZED NEO-FUZZY NEURONS AND ITS LEARNING BASED ON BACKPROPAGATION

Y. Bodyanskiy[1], T. Antonenko[2]
[1,2]Kharkiv National University of Radioelectronics, Ukraine
Nauky Ave, 14, Kharkiv, 61166

[1]http://orcid.org/0000-0001-5418-2143
[2]http://orcid.org/0000-0002-9421-707X

Abstract. Modern approaches in deep neural networks have a number of issues related to the learning process and computational costs. This article considers the architecture grounded on an alternative approach to the basic unit of the neural network. This approach achieves optimization in the calculations and gives rise to an alternative way to solve the problems of the vanishing and exploding gradient.

The main issue of the article is the usage of the deep stacked neo-fuzzy system, which uses a generalized neo-fuzzy neuron to optimize the learning process. This approach is non-standard from a theoretical point of view, so the paper presents the necessary mathematical calculations and describes all the intricacies of using this architecture from a practical point of view.

From a theoretical point, the network learning process is fully disclosed. Derived all necessary calculations for the use of the backpropagation algorithm for network training.

A feature of the network is the rapid calculation of the derivative for the activation functions of neurons. This is achieved through the use of fuzzy membership functions. The paper shows that the derivative of such function is a constant, and this is a reason for the statement of increasing in the optimization rate in comparison with neural networks which use neurons with more common activation functions (ReLU, sigmoid).

The paper highlights the main points that can be improved in further theoretical developments on this topic. In general, these issues are related to the calculation of the activation function. The proposed methods cope with these points and allow approximation using the network, but the authors already have theoretical justifications for improving the speed and approximation properties of the network.

The results of the comparison of the proposed network with standard neural network architectures are shown.

Keywords: deep stacking neural network, neo-fuzzy neuron, multilayer neural network, F-transform.

ГЛИБОКА НЕЙРОННА МЕРЕЖА НА ОСНОВІ УЗАГАЛЬНЕНИХ НОВОНЕЧІТКИХ НЕЙРОНІВ ТА ЇЇ НАВЧАННЯ ЗА ДОПОМОГОЮ ЗВОРОТНОГО ПОШИРЕННЯ ПОХИБКИ

Є.В. Бодянський[1], Т.Є. Антоненко[2]
[1]Харківський національний університет радіоелектроніки, Україна
пр. Науки, 14, м.Харків, 61166

Анотація. У сучасних підходах до глибоких нейронних мереж виникає ряд актуальних питань, пов'язаних із процесом навчання та з обчислювальними затратами. У статті розглянуто архітектуру нейронної мережі, у якій реалізовано альтернативний підхід до базової одиниці нейронної мережі. За рахунок цього досягається оптимізація обчислень і з'являється новий погляд на розв'язання відомих проблем глибоких мереж – зникального та вибухального градієнта.

У статті розглянуто глибоку стекову нову нечітку систему, в якій використано узагальнений ново-нечіткий нейрон для оптимізації процесу навчання. З теоретичного погляду такий підхід є нестандартним, тож у роботі наведено необхідні математичні викладки та описано всі практичні тонкощі використання цієї архітектури.

З теоретичної сторони повністю розкрито процес навчання такої мережі. Зроблені всі необхідні викладки щодо використання алгоритму зворотного поширення похибки для навчання цієї мережі.

Особливістю мережі є швидке обчислення похідної для активаційних функцій нейронів. Це досягається за рахунок використання нечітких (фаззі) функцій належності. В роботі показано, що похідна такої функції є

константою, а це є приводом для того, щоб припустити наявність приросту за швидкістю оптимізації у порівнянні з нейронними мережами, що використовують нейрони з більш поширеними функціями активації (ReLU, sigmoid).

У роботі висвітлено основні аспекти, які можна покращити у подальших теоретичних дослідженнях на цю тему. Загалом ці питання пов'язані з обчисленням функції активації. Запропоновані методи справляються з цим завданням і дозволяють проводити апроксимацію за допомогою мережі, але автори вже мають теоретичні обґрунтування для покращення швидкодії та апроксимаційних властивостей мережі.

У роботі показані результати порівняння запропонованої конфігурації зі стандартними архітектурами нейронних мереж.

Ключові слова: глибока стекова мережа, новий нечіткий нейрон, багатошарова нейронна мережа, F-перетворення.

Вступ

На сьогодні глибокі нейронні мережі (ГНМ) широко застосовують для розв'язання різноманітних завдань з видобування даних, включаючи опрацювання зображень і текстів, прогнозування та діагностику, інтелектуальне керування та прийняття рішень тощо [1-5], забезпечуючи при цьому високу якість отриманих результатів. У той же час слід відзначити і деяку громіздкість цих систем, що призводить до суттєвого зниження їх швидкодії та породжує низку обчислювальних проблем у процесі навчання. Основним «будівельним блоком» ГНМ є елементарний перцептрон Розенблатта з активаційними функціями (зазвичай типу ReLU), що не задовольняють вимоги основних апроксимаційних теорем, які лежать в основі традиційної теорії штучних нейронних мереж (ШНМ) – у сучасній термінології мілких (*англ.* *shallow*) нейронних мереж (МНМ). У зв'язку із цим, у [6] було запропоновано замість перцептронів Розенблатта використовувати ново-нечіткі нейрони (ННН) [7-9], що є за своєю суттю досить простими ново-нечіткими системами типу Такагі-Сугено-Канга нульового порядку, тобто мають універсальні апроксимаційні властивості. Запропонована глибока ново-нечітка нейронна мережа забезпечила високу якість опрацювання інформації, не страждаючи в процесі навчання від проблем «зникального» або «вибухального» градієнта. До недоліків цієї системи слід віднести істотне збільшення кількості функцій належності, що використовуються у нових нечітких нейронах замість традиційних функцій активації у перцептроні Розенблатта.

Цю проблему можна подолати, скориставшись ідеєю стекових нейронних мереж [2, 10, 11], де шари утворюються не окремими нейронами, а нейро-мережевими блоками (СНН). Як такі блоки-стеки, доцільно використати так звані узагальнені нові нечіткі нейрони (УННН) [12], що відрізняються наявністю декількох виходів замість одного у звичайного ННН. На основі УННН у [13, 14] було запропоновано двошарові гібридні системи обчислювального інтелекту, а в [15, 16] – глибокі мережі. Останні вже довели свою ефективність, проте зростання кількості входів у кожному наступному стеку-каскаді робить ці системи занадто громіздкими. У зв'язку із цим, видається доцільним розглянути глибоку нейронну мережу з прямою передачею інформації (feedforward architecture) із системами типу узагальнених нових нечітких нейронів і чисельно стійкий алгоритм її навчання на основі зворотного поширення похибки.

1. Архітектура глибокої нейронної мережі на основі узагальнених нових нечітких нейронів

На рис. 1 наведено архітектуру (feedforward architecture) багатошарової стекової нової нечіткої нейронної мережі (БСНННМ), що містить s шарів-стеків. На вхід мережі (нульовий шар) надходить вхідний вектор образів $x(k) = \left(x_1(k), x_2(k), \dots, x_n(k)\right) \epsilon\, R^n$ (тут k=1, 2, …, N – номер спостереження у навчальній вибірці або індекс поточного дискретного часу), а на виході мережі (вихідного s-го шару) формується вихідний векторний сигнал.

$$\hat{y}(k) = \left(\hat{y}_1(k), \hat{y}_2(k), \dots, \hat{y}_m(k)\right)^T \in R^m$$

Входом p-го шару ($p=1, 2\dots, s$) є вихід попереднього $(p-1)$-го шару

$$o^{[p-1]}(k) = (o_1^{[p-1]}(k), \dots, o_{i_{p-1}}^{[p-1]}(k), \dots, o_{n_{p-1}}^{[p-1]}(k))^T \in R^{n_{p-1}},$$

а його виходом – векторний сигнал.

$$o^{[p]}(k) = (o_1^{[p]}(k), \dots, o_{i_p}^{[p]}(k), \dots, o_{n_p}^{[p]}(k))^T \in R^{n_p}$$

Далі для зручності математичних перетворень будемо також використовувати позначення

$$x(k) \equiv o^{[0]}(k) = (o_1^{[0]}(k), \dots, o_{i_0}^{[0]}(k) \ \dots, o_{n_0}^{[0]}(k))^T$$

та

$$\hat{y}(k) \equiv o^{[s]}(k) = (o_1^{[s]}(k), \dots, o_{i_s}^{[s]}(k), \dots, o_{n_s}^{[s]}(k))^T, n \equiv n_0, m \equiv n_s.$$

Таким чином, БСНННМ реалізує нелінійне відображення $R^n \to R^m$, параметри якого відновлюються в процесі навчання нейронної мережі.

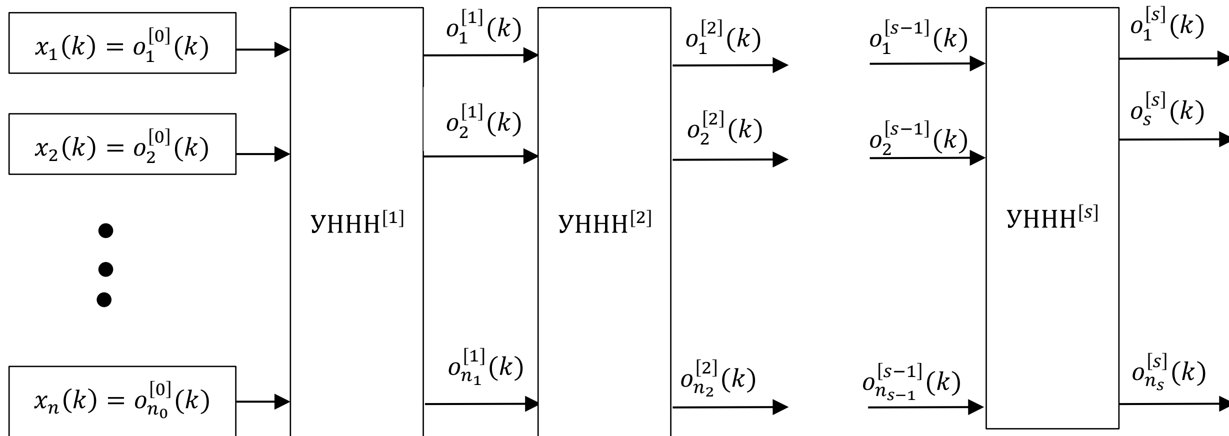


Рис 1. Архітектура багатошарової стекової нової нечіткої нейронної мережі

2.Еволюційний узагальнений новий нечіткий нейрон як стек-шар глибокої нейронної мережі

На рис. 2 наведено схему узагальненого нового нечіткого нейрона УННН ^{p}, який утворює p-й шар (включаючи перший прихований $p=1$ і вихідний шари $p=s$) глибокої нейронної мережі. Кожен такий нейрон містить n_{p-1} паралельних нелінійних синапсів БНС$_{i_{p-1}}^{[p]}$, кожен з яких має один вхід і, на відміну від традиційних нелінійних синапсів [7-9], n_p виходів. Кожен багатовимірний нелінійний синапс містить h функцій належності $\mu_{li_{p-1}}^{[p]}(O_{i_{p-1}}^{p-1})$ ($l = 1, 2, \dots, h$) і $n_p h$ налаштованих синапсів синаптичних ваг $w_{i_p l i_{p-1}}^{[p]}$, які уточнюються в процесі навчання. Таким чином, кожен шар містить hn_{p-1} функцій належності та $n_p h n_{p-1}$ синаптичних ваг.

Нелінійне відображення, що реалізується УННН,^{p} визначається типом функцій належності, що використовуються у конкретній мережі, та може бути записано у вигляді

$$o_{i_p}^{[p]}(k) = \sum_{i_{p-1}=1}^{n_{p-1}} \sum_{l=1}^{h} w_{i_p l i_{p-1}}^{[p]} \mu_{l i_{p-1}}^{[p]} \left(o_{i_{p-1}}^{[p-1]}(k)\right), \forall i_p = 1, 2, \dots, n_p. \quad (1)$$

Відзначимо також, що для вихідного шару мережі, вираз (1) може бути записаний у формі

$$y_j(k) \equiv o_{i_s}^{[s]}(k) = \sum_{i_{s-1}=1}^{n_{s-1}} \sum_{l=1}^{h} w_{i_s l i_{s-1}}^{[s]} \mu_{l i_{s-1}}^{[s]} \left(o_{i_{s-1}}^{[s-1]}(k)\right), i_p = 1,2,\dots,m$$

Як функції належності, у найпростішому випадку можуть бути використані трикутні конструкції.

$$\mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \begin{cases} \dfrac{o_{i_{p-1}}^{[p-1]} - c_{l-1,i_{p-1}}^{[p]}}{c_{li_{p-1}}^{[p]} - c_{l-1,i_{p-1}}^{[p]}}, \text{якшо } o_{i_{p-1}}^{[p-1]} \in \left[c_{l-1,i_{p-1}}^{[p]}, c_{li_{p-1}}^{[p]}\right], \\ \dfrac{c_{l+1,i_{p-1}}^{[p]} - o_{i_{p-1}}^{[p-1]}}{c_{l+1,i_{p-1}}^{[p]} - c_{li_{p-1}}^{[p]}}, \text{якщо } o_{i_{p-1}}^{[p-1]} \in \left[c_{l,i_{p-1}}^{[p]}, c_{l+1,i_{p-1}}^{[p]}\right], \\ 0, \quad \text{інакше}, \end{cases}$$

$$c_{1i_{p-1}}^{[p]} = o_{i_{p-1,min}}^{[p-1]}, c_{hi_{p-1}}^{[p]} = o_{i_{p-1,max}}^{[p-1]},$$

що задовольняють умови одиничного розбиття Руспіні:

$$\begin{cases} \mu_{l-1,i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) + \mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = 1, \text{якшо } o_{i_{p-1}}^{[p-1]} \in \left[c_{l-1,i_{p-1}}^{[p]}, c_{li_{p-1}}^{[p]}\right], \\ \mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) + \mu_{l+1,i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = 1, \text{якшо } o_{i_{p-1}}^{[p-1]} \in \left[c_{li_{p-1}}^{[p]}, c_{l+1,i_{p-1}}^{[p]}\right], \end{cases}$$

де $c_{li_{p-1}}^{[p]}, l = 1, 2, \dots, h$ - центри відповідних функцій належності,

$$o_{i_{p-1,min}}^{[p-1]}, o_{i_{p-1,max}}^{[p-1]} -$$

мінімальне та максимальне значення вхідного сигналу i_{p-1}-го багатовимірного нелінійного синапсу БНС$_{i_{p-1}}^{[p]}$ стеку УННН ^{p}.

Таким чином, якщо центри активаційних функцій $\mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)$ рівномірно розподілені на інтервалі $\left[o_{i_{p-1,min}}^{[p-1]}, o_{i_{p-1,max}}^{[p-1]}\right]$, відстань між двома сусідніми центами визначається величиною

$$\Delta^{[p]} = \frac{o_{i_{p-1,max}}^{[p-1]} - o_{i_{p-1,min}}^{[p-1]}}{h-1}.$$

У реальних задачах, проте, вихідні сигнали $o_{i_{p-1}}^{[p-1]}(k)$ можуть виходити за означений інтервал і в цьому випадку конкретний БНС$_{i_{p-1}}^{[p]}$ може просто не реагувати на подане на його вхід збудження. Подолати це ускладнення можна, скориставшись ідеєю еволюційного нового нечіткого нейрона [19], який у нашому випадку набуває досить простої форми.

Отже, нехай на вхід БНС$_{i_{p-1}}^{[p]}$ надійшло значення сигналу $o_{i_{p-1}}^{[p-1]*}(k) > o_{i_{p-1,max}}^{[p-1]}$, яке виходить понад заданий інтервал. У цьому випадку функція належності $\mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)$ доповнюється справа компонентою

$$\mu_{hi_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \frac{o_{i_{p-1}}^{[p-1]*}(k) - o_{i_{p-1}}^{[p-1]}(k)}{o_{i_{p-1}}^{[p-1]*}(k) - c_{h,i_{p-1}}^{[p]}}$$

для

$$o_{i_{p-1}}^{[p-1]} \in [c_{h,i_{p-1}}^{[p]} = o_{i_{p-1,max}}^{[p-1]}, o_{i_{p-1}}^{[p-1]*}(k)].$$

Крім того, формується нова функція належності

$$\mu_{h+1,i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \frac{o_{i_{p-1}}^{[p-1]}(k) - c_{h,i_{p-1}}^{[p]}}{o_{i_{p-1}}^{[p-1]*}(k) - c_{h,i_{p-1}}^{[p]}}$$

для

$$o_{i_{p-1}}^{[p-1]} \in [c_{h,i_{p-1}}^{[p]} = o_{i_{p-1,max}}^{[p-1]}, o_{i_{p-1}}^{[p-1]*}(k)].$$

У протилежному випадку, коли на вхід надходить значення $o_{i_{p-1^*}}^{[p-1]}(k) < o_{i_{p-1,min}}^{[p-1]}$, функція $\mu_{1,i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right)$ доповнюється зліва компонентою

$$\mu_{1i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \frac{o_{i_{p-1^*}}^{[p-1]}(k) - o_{i_{p-1}}^{[p-1]}(k)}{c_{1,i_{p-1}}^{[p]} - o_{i_{p-1^*}}^{[p-1]}(k)},$$

для

$$o_{i_{p-1}}^{[p-1]} \in [o_{i_{p-1^*}}^{[p-1]}(k), c_{1,i_{p-1}}^{[p]} = o_{i_{p-1,min}}^{[p-1]}].$$

Також формується нова функція належності з центром

$$c_{0,i_{p-1}}^{[p]} = o_{i_{p-1*}}^{[p-1]}(k):$$

$$\mu_{0i_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right) = \frac{c_{1,i_{p-1}}^{[p]} - o_{i_{p-1}}^{[p-1]}(k)}{c_{1,i_{p-1}}^{[p]} - o_{i_{p-1*}}^{[p-1]}(k)},$$

для

$$o_{i_{p-1}}^{[p-1]} \epsilon \left[o_{i_{p-1*}}^{[p-1]}(k), c_{1,i_{p-1}}^{[p]} = o_{i_{p-1,min}}^{[p-1]}\right]$$

Таким чином, багатовимірний нелінійний синапс $БНС_{i_{p-1}}^{[p]}$ набуває здатності опрацьовувати сигнали в діапазоні

$$o_{i_{p-1*}}^{[p-1]} \leq o_{i_{p-1}}^{[p-1]} \leq o_{i_{p-1}}^{[p-1]*}.$$

Якщо ж у процесі навчання виникає ситуація, коли вхідний сигнал $o_{i_{p-1}}^{[p-1]}(k)$ виходить із цього діапазону, процедура еволюції системи функцій належності може аналогічним чином бути продовжена.

3.Навчання глибокої нейронної мережі на основі узагальнених нових нечітких нейронів

Процес навчання БСНННМ реалізується на основі зворотного поширення похибки і є за суттю градієнтною процедурою мінімізації прийнятого критерію навчання, за який найзручніше прийняти квадратичну функцію.

$$E(k) = \frac{1}{2}\sum_{i_s=1}^{n_s} e_{i_s}^2(k) = \frac{1}{2}\sum_{j=1}^{m} e_j^2(k), \qquad (2)$$

де

$e_{i_s}(k) \equiv e_j(k) = y_{i_s}(k) - \sum_{i_{s-1}=1}^{n_{s-1}}\sum_{l=1}^{h} w_{i_s l i_{s-1}}^{[s]} \mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right) \equiv y_j(k) - \sum_{i_{s-1}=1}^{n_{s-1}}\sum_{l=1}^{h} w_{j l i_{s-1}}^{[s]} \mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right), y_{i_s}(k) \equiv y_j(k)$ - зовнішній навчальний сигнал.

Вводячи до розгляду вектор навчальних сигналів

$$y(k) = (y_1(k), y_2(k), \dots, y_j(k), \dots y_m(k))^T \equiv (y_1(k), \dots, y_{i_s}(k), \dots y_{n_s}(k))^T,$$

$(hn_s \times 1)$ – вектор функцій належності вихідного шару.

$$\mu^s\left(o^{[s-1]}(k)\right) = \left(\mu_{1,1}^{[s]}\left(o_1^{[s-1]}(k)\right), \dots, \mu_{h,1}^{[s]}\left(o_1^{[s-1]}(k)\right), \dots, \right.$$
$$\left., \dots, \mu_{l,i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right), \dots, \mu_{h,n_{s-1}}^{[s]}\left(o_{n_{s-1}}^{[s-1]}(k)\right)\right)^T$$

та $(n_s \times hn_{s-1})$ – матрицю синаптичних ваг вихідного шару $W^s = \left\{w_{i_s l i_{s-1}}^{[s]}\right\}$, можна переписати критерій навчання у більш компактній формі

$$E(k) = \frac{1}{2}\left\|y - W^s \mu^{[s]}\left(o^{[s-1]}(k)\right)\right\|^2 = \frac{1}{2}\|e(k)\|^2. \qquad (3)$$

Аналогічним чином можна також записати нелінійне відображення, що реалізується кожним шаром-стеком мережі:

$$o^{[p]}(k) = W^p \mu^{[p]}\left(o^{[p-1]}(k)\right), \forall p = 1,2, \dots s,$$

де $W^p = \left\{w_{i_p l i_{p-1}}^{[p]}\right\} - \left(n_p \times hn_{p-1}\right)$ – матриця синаптичних ваг p-го шару, що підлягає визначенню у процесі навчання.

Для налаштування вихідного шару може бути використаний матричний алгоритм навчання, що має як слідкуючі, так і згладжуючі властивості, у вигляді [14].

$$\begin{cases} W^{[s]}(k) = W^{[s]}(k-1) + \eta(k)e(k)\mu^{[s]T}\left(o^{[s-1]}(k)\right) = \\ = W^{[s]}(k-1) + \left(r^{[s]}(k)\right)^{-1} e(k)\mu^{[s]T}\left(o^{[s-1]}(k)\right) = \\ = W^{[s]}(k-1) + \left(r^{[s]}(k)\right)^{-1} * \\ * \left(y(k) - W^{[s]}(k-1)\mu^{[s]}\left(o^{[s-1]}(k)\right)\right) * \\ * \mu^{[s]T}\left(o^{[s-1]}(k)\right) \\ r^{[s]}(k) = \alpha r^{[s]}(k-1) + \left\|\mu^{[s]}\left(o^{[s-1]}(k)\right)\right\|^2, \end{cases} \qquad (4)$$

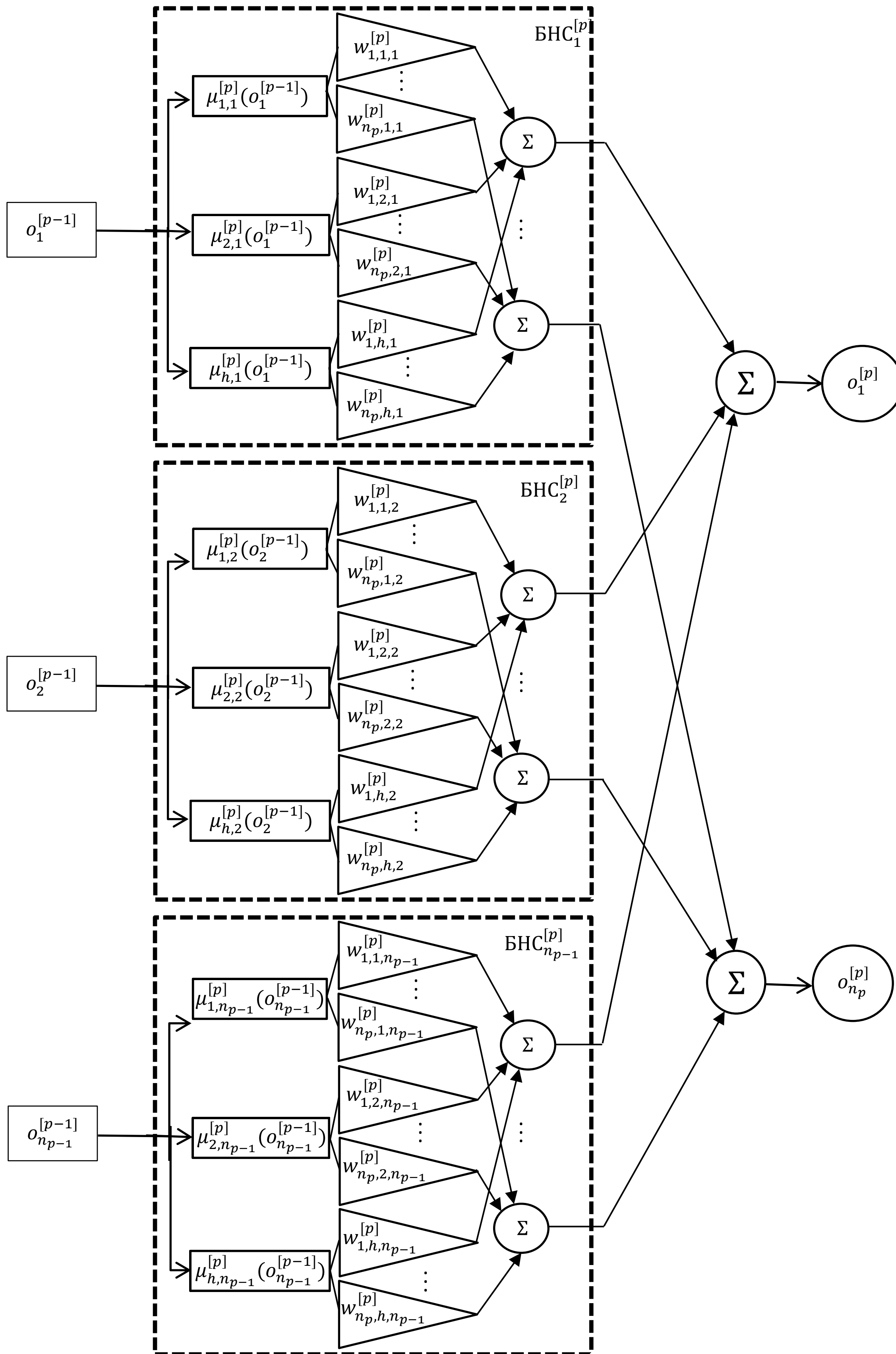


Рис 2. Узагальнений новий нечіткий нейрон *p*-го шару багатошарової стекової нової нечіткої нейронної мережі БСНННМ

де $0 \leq \alpha \leq 1$ – фактор забування, $\eta(k)$ – параметр кроку навчання, що при $\alpha = 0$ набуває матричну форму оптимального за швидкодією алгоритму навчання Качмажа-Уідроу-Хоффа (Kaczmarz-Widrow-Hoff)

$$W^{[s]}(k) = W^{[s]}(k-1) + \\ + \frac{e(k)\mu^{[s]T}\left(o^{[s-1]}(k)\right)}{\mu^{[s]T}\left(o^{[s-1]}(k)\right) * \mu^{[s]}\left(o^{[s-1]}(k)\right)} = W^{[s]}(k - \\ 1) + + e(k)\mu^{[s]+}\left(o^{[s-1]}(k)\right), \quad (5)$$

де $(\cdot)^{+}$ - символ псевдообернення за Муром-Пенроузом. Можна довести, що алгоритм (5) є найшвидкішим в класі градієнтних алгоритмів навчання. Нескладно також помітити, що при $\alpha = 1$ (4) перетворюється у процедуру стохастичної апроксимації.

Вводячи до розгляду i_s -й рядок матриці $W^{[s]}$,

$$W_{i_s}^{[s]} = (W_{i_s 11}^{[s]}, \dots, W_{i_s h1}^{[s]}, W_{i_s 12}^{[s]}, \dots, W_{i_s h n_{s-1}}^{[s]})$$

розмірності $(1 \times hn_{s-1})$, на основі (5), нескладно ввести процедуру налаштування синаптичних ваг $i_s = j$-го виходу системи

$$W_{i_s}^{[s]}(k) = W_{i_s}^{[s]}(k-1) + \\ + \frac{\left(y_{i_s}(k) - W_{i_s}^{[s]}(k-1) * \mu^{[s]}\left(o^{[s-1]}(k)\right)\right)}{\mu^{[s]T}\left(o^{[s-1]}(k)\right) * \mu^{[s]}\left(o^{[s-1]}(k)\right)} \mu^{[s]T}\left(o^{[s-1]}(k)\right) = \\ W_j^{[s]}(k-1) +$$

$$\frac{\left(y_j(k) - W_j^{[s]}(k-1) * \mu^{[s]}\left(o^{[s-1]}(k)\right)\right)}{\left\| * \mu^{[s]}\left(o^{[s-1]}(k)\right) \right\|^2} \mu^{[s]T}\left(o^{[s-1]}(k)\right) = \\ W_j^{[s]}(k-1) + e_j(k)\mu^{[s]+}\left(o^{[s-1]}(k)\right).$$

І, нарешті, можна записати алгоритм навчання окремі $i_s l i_{s-1}$-ї ваги

$$w_{i_s l i_{s-1}}^{[s]} \\ = w_{i_s l i_{s-1}}^{[s]}(k-1) \\ + \frac{e_{i_s}(k)\mu_{l i_{s-1}}{}^{[s]T}\left(o_{i_{s-1}}{}^{[s-1]}(k)\right)}{\sum_{i_{s-1}=1}^{n_{s-1}} \sum_{l=1}^{h} (\mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right))^2}$$

або у більш загальній формі

$$w_{i_s l i_{s-1}}^{[s]}(k) = w_{i_s l i_{s-1}}^{[s]}(k-1) + \\ + \eta(k) * \\ * \left(y_{i_s}(k) - \sum_{i_{s-1}=1}^{n_{s-1}} \sum_{l=1}^{h} w_{i_s l i_{s-1}}^{[s]}(k \\ - 1)\mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right) \right) * \\ * \mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right), \quad (6)$$

швидкість збіжності якої визначається вибором параметру кроку $\eta(k)$.

Налаштування синаптичних ваг прихованих шарів від *(s-1)*-го до першого відбувається за допомогою процедури зворотного поширення похибки, для чого може бути використаний алгоритм типу (6) у вигляді (для передостаннього *(s-1)*-го шару):

$$w_{i_{s-1} l i_{s-2}}^{[s-1]}(k) = w_{i_{s-1} l i_{s-2}}^{[s-1]}(k - \\ 1) - \\ - \eta(k) * \frac{\partial \mathrm{E}(k)}{\partial w_{i_{s-1} l i_{s-2}}^{[s-1]}} = \\ = w_{i_{s-1} l i_{s-2}}^{[s-1]}(k-1) - \eta(k) * \\ \frac{\partial (e_{i_s}(k))^2}{\partial o_{i_s}^{[s]}(k)} * \frac{\partial o_{i_s}^{[s]}(k)}{\partial o_{i_{s-1}}^{[s-1]}(k)} * \frac{\partial o_{i_{s-1}}^{[s-1]}(k)}{\partial w_{i_{s-1} l i_{s-2}}^{[s-1]}}. \quad (7)$$

З урахуванням того, що

$$\frac{\partial o_{i_s}^{[s]}(k)}{\partial o_{i_{s-1}}^{[s-1]}(k)} = \sum_{l=1}^{h} w_{i_s l i_{s-1}}^{[s]} \frac{\mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right)}{\partial o_{i_{s-1}}^{[s-1]}(k)},$$

де

$$\frac{\mu_{l i_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right)}{\partial o_{i_{s-1}}^{[s-1]}(k)} \\ = \begin{cases} (c_{l i_{s-1}}^{[s]} - c_{l-1, i_{s-1}}^{[s]})^{-1}, \text{якщо } o_{i_{s-1}}^{[s-1]} \in [c_{l-1, i_{s-1}}^{[s]}, c_{l i_{s-1}}^{[s]}] \\ -(c_{l+1, i_{s-1}}^{[s]} - c_{l, i_{s-1}}^{[s]})^{-1}, \text{якщо } o_{i_{s-1}}^{[s-1]} \in [c_{l, i_{s-1}}^{[s]}, c_{l+1, i_{s-1}}^{[s]}] \\ 0 \text{ інакше,} \end{cases}$$

алгоритм (7) остаточно можна записати у вигляді

$$w_{i_{s-1}li_{s-2}}^{[s-1]}(k) = w_{i_{s-1}li_{s-2}}^{[s-1]}(k-1) +$$
$$+\eta(k)\sum_{i_s=1}^{n_s} e_{i_s}(k) \quad (8)$$
$$*\sum_{l=1}^{h} w_{i_{s-1}li_{s-2}}^{[s-1]} \frac{\mu_{li_{s-1}}^{[s]}\left(o_{i_{s-1}}^{[s-1]}(k)\right)}{\partial o_{i_{s-1}}^{[s-1]}(k)} \mu_{li_{s-2}}^{[s-1]}\left(o_{i_{s-2}}^{[s-2]}(k)\right).$$

Процедура налаштування прихованих шарів може бути записана аналогічним чином. При цьому

$$\frac{\partial \mathrm{E}(k)}{\partial o_{i_p}^{[p]}} = \sum_{i_{p+1}=1}^{n_{p+1}} \frac{\partial \mathrm{E}}{\partial o_{i_{p+1}}^{[p+1]}} * \frac{\partial o_{i_{p+1}}^{[p+1]}}{\partial o_{i_p}^{[p]}},$$

$$\delta_{i_p}^{[p]} = \sum_{i_{p+1}=1}^{n_{p+1}} \delta_{i_{p+1}}^{[p+1]} * \frac{\partial o_{i_{p+1}}^{[p+1]}}{\partial o_{i_p}^{[p]}},$$

(тут $\delta_{i_p}^{[p]}$ - δ-похибка i_p-го шару)

$$\frac{\partial o_{i_{p+1}}^{[p+1]}}{\partial o_{i_p}^{[p]}} = \sum_{l=1}^{h} w_{i_{p+1}li_p}^{[p+1]} \frac{\mu_{li_p}^{[p+1]}\left(o_{i_p}^{[p]}\right)}{\partial o_{i_p}^{[p]}},$$

$$\frac{\partial o_{i_p}^{[p]}}{\partial w_{i_pli_{p-1}}^{[p]}} = \mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}\right),$$

а кінцевий результат набуває форми

$$w_{i_pli_{p-1}}^{[p]}(k) = w_{i_pli_{p-1}}^{[p]}(k-1)$$
$$-\eta(k)\frac{\partial \mathrm{E}(k)}{\partial w_{i_pli_{p-1}}^{[p]}} = w_{i_pli_{p-1}}^{[p]}(k-1)$$
$$+\eta(k)\sum_{i_{p+1}=1}^{n_{p+1}} \delta_{i_{p+1}}^{[p+1]} * \frac{\partial o_{i_{p+1}}^{[p+1]}}{\partial o_{i_p}^{[p]}} \quad (9)$$
$$* \mu_{li_{p-1}}^{[p]}\left(o_{i_{p-1}}^{[p-1]}(k)\right).$$

Завдяки тому, що похідні функцій належності є сталими, навчання розглядуваної системи є досить простим з обчислювальної точки зору, а універсальні апроксимуючі властивості шарів-стеків, що є у загальному випадку нейро-нечіткими системами Такагі-Сугено-Канга, забезпечують високу якість опрацювання інформації.

4. Експериментальні дослідження

У цьому експерименті було проведено порівняння результатів навчання на якість апроксимації багатошарового перцептрона з сигмоїдними функціями активації та запропонованої глибокої нової нечіткої мережі. Задля тесту використовувався датасет «Breast Cancer Wisconsin Data Set (BC dataset)».

Завданням експерименту було з’ясувати, чи може запропонована модель показувати у порівнянні з популярними моделями результати і чи можна її використовувати як альтернативну.

Таблиця 1. Порівняння результатів навчання БШП та БСНННМ

Тип мережі	БШП		БСНННМ		
Архітектура	3Х100	6Х100	3Х10	6Х10	
Точність на 7-ій епосі (тест)	92.1	93.85	93.86	91.23	
Функція втрат на 7-ій епосі (тест)	0.163	0.142	0.41	0.412	
Функція втрат на 7-ій епосі (тренування)	0.179	0.123	0.429	0.397	

Цікавим є те, що у даному експерименті на 7-ій епосі БСНННМ із шести шарів продемонструвала нижчу точність, ніж варіант з трьома шарами. Разом з тим, результати функції втрат показують приріст в оптимальності. Такі результати демонструють елемент стохастичності при навчанні мережі.

Висновки

У статті запропоновано глибоку стекову нову нечітку нейронну мережу, шари якої утворені узагальненими новими нечіткими нейронами, при цьому кількість функцій належності у кожному шарі може змінюватись у процесі навчання. Така мережа забезпечує високі апроксимаційні властивості, оскільки кожен стек є за своєю

суттю нейр-онечіткою системою Такагі-Сугено-Канга, при цьому вихідний сигнал мережі та стеків лінійно залежить від налаштованих синаптичних ваг, що забезпечує високу швидкість навчання. Крім того, запропонована система є досить простою і швидкою з погляду обчислювальної реалізації.

References

1. Bengio Y, LeCun Y, Hinton G. Deep Learning – Nature – 2015-521 – p.436-444.
2. Schmidhuber J Deep learning in neural networks: An overview – Neural Networks – 2015–01 – p.85-117.
3. Googfellow I, Bengio Y, Courville A. Deep Learning – MIT Press, 2016-787p.
4. Graupe D. Deep Learning Neural Networks: Design and Case Studies- New York: World Scientific, 2016 – 260p.
5. Caterini A.L., Chang D.E. Deep Neural Networks in a Mathematical Framework – Springer, 2018 –79p.
6. Cichocki A, Unbehauen R. Neural Networks for Optimization and Signal Processing – Stuttgart: Teubner, 1993-526p.
7. Cybenko G. Approximation by superpositions of a sigmoidal function – Math. Control Signals Systems. – 1985 – 2 – p.303-314.
8. Hornik K. Approximation capabilities of multilayer feedforward networks – Neural Networks, - 1994 – 4 – p.251-257.
9. Aggarwal Ch.C. Neural Networks and Deep Learning –Springer, 2018-512p.
10. Yamakawa T, Uchino E, Miki T., Kusanagi H. A neo fuzzy neuron and its applications to system identification and predictions to system behavior. – Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks, pp. 477-483, 1992.
11. Uchino E, Yamakava T. Neo-fuzzy neuron based new approach to system modeling with application to actual system - Proceedings Sixth International Conference on Tools with Artificial Intelligence – New Orlean, LA, USA, 1994 – p.564-570.
12. Miki T, Yamakawa T, “Analog implementation of neo-fuzzy neuron and its on-board learning,” In Computational Intelligence and Applications, Piraeus: WSES Press, 1999, pp. 144-149.
13. Kolodyazhniy V, Bodyanskiy Ye. Fuzzy Kolmogorov's network – Lecture Notes in Computer Science. – 3214 – Heidelberg: Springer Verlag, 2004. – p.764-771.
14. Bodyanskiy Ye, Kolodyazhniy V, Otto P. Neuro-fuzzy Kolmogorov's network for time series prediction and pattern classification – Lecture Notes in Artificial Intelligence – 3698 – Heidelberg: Springer Verlog, 2005. – p.191-202.
15. Bodyanskiy Ye,Popov S, Rybalchenko T. Multilayer neuro-fuzzy network for short term electric load forecasting – Lecture Notes in Computer Science. – 5010 – Berlin-Heidelberg: Springer Verlag, 2008. – p.339-348.
16. Bodyanskiy Ye,Vynokurova O, Setlak G, Peleshko D, Mulesa P. Adaptive multivariate hybrid neuro-fuzzy system and its on-board fast learning – Neurocomputing – 2017 – 230-p.409-416.
17. Perfilieva I. Fuzzy transforms: Theory and applications – Fuzzy Sets and Systems – 2006 – 157 – p.993-1023.
18. Bodyanskiy Ye, Kolodyazhniy V, Stephan A. An adaptive learning algorithm for a neuro-fuzzy network - Ed. by B.Reush “Computitional Intelligence. Theory and Application” – Berlin-Heidelberg: Ney York: Springer, 2001. – p.68-75.
19. Otto P, Bodyanskiy Ye, Kolodyazhniy V. A new learning algorithm for a forecasting neuro-fuzzy network - Integrated Computer Aided Engineering – 2003 – 10(4) – p.399-409.

Література

1. Deep Learning / Бенжіо Й., Лікун Я., Хінтон Дж. // Nature: Науковий журнал. – 2015. – № 521 – С.436-444.
2. Deep learning in neural networks: An overview / Шмідхубер Ю. // Neural Networks: Науковий журнал. – 2015–01 С.85-117.
3. Deep Learning / [Я. Гудфеллоу, Й. Бенжіо, А. Коурвіль]; – MIT Press, 2016. -787с.
4. Грауп Д. Deep Learning Neural Networks: Design and Case Studies / New York: World Scientific, 2016. – 260с.
5. Deep Neural Networks in a Mathematical Framework / [Катерині А., Чанг Д.]; – Springer, 2018. –79с.
6. Neural Networks for Optimization and Signal Processing / [Чічокі А., Унбехан Р.]; – Stuttgart: Teubner, 1993. -526с.
7. Цибенко Дж. Approximation by superpositions of a sigmoidal function /– Math. Control Signals Systems. – 1985 – 2 – с.303-314.
8. Горнік К. Approximation capabilities of multilayer feedforward networks – 1994 – 4 – с.251-257.
9. Аггарвал Ч. Neural Networks and Deep Learning – Cham: Springer, 2018-512с.
10. Нео фаззі нейрон та його використання для систем ідентифікації. / [Ямакава Т., Учино І., Мікі Т., Кусанагі Х.,] Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks, - 1992. - с.477-483.
11. Учино І., Ямакава Т., / A neo fuzzy neuron and its applications to system identification and predictions to system behavior. - Proceedings Sixth International Conference on Tools with Artificial Intelligence – New Orlean, LA, USA, 1994 – с.564-570.
12. Мікі Т., Ямакава Т., / Analog implementation of neo-fuzzy neuron and its on-board learning. In Computational Intelligence and Applications, Piraeus: WSES Press, 1999, с.144-149.
13. Колодязний В., Бодянській Є.В., / Fuzzy Kolmogorov's network – Лекції ШІ. – 3214 – Heidelberg: Springer Verlag, 2004. – с.764-771.
14. Бодянській Є.В., Колодязний В., Отто П. / Neuro-fuzzy Kolmogorov's network for time series prediction and pattern classification – Лекції ШІ –

3698 – Heidelberg: Springer Verlog, 2005. – с.191-202.
15. Бодянській Є.В., Попов С., Рибалченко Т., / Multilayer neuro-fuzzy network for short term electric load forecasting – Лекції ШІ. – 5010 – Berlin-Heidelberg: Springer Verlag, 2008. – с.339-348.
16. Бодянській Є.В., Виконкурова О., Селтак Дж., Пелешко Д., Мулеса П., Adaptive multivariate hybrid neuro-fuzzy system and its on-board fast learning – Neurocomputing – 2017 – 230-с.409-416.
17. Перфільєва І., / Fuzzy transforms: Theory and applications // – Нечіткі множини та системи – 2006 – 157 – с.993-1023.
18. Бодянській Є.В., Колодязний В., Стефан А./ An adaptive learning algorithm for a neuro-fuzzy network - Ed. by B.Reush “Computitional Intelligence. Theory and Application” – Berlin-Heidelberg: Ney York: Springer, 2001. – с.68-75.
19. Отто П., Бодянській Є.В., Колодязний В., / A new learning algorithm for a forecasting neuro-fuzzy network - Integrated Computer Aided Engineering – 2003 – 10(4) – с.399-409.

Received 15.04.21
Accepted 01.06.21